

Supplementary Box 2. Lambda coalescents

Most software for performing demographic inference from population genomics data assumes that the data were generated under a Kingman coalescent model⁷. While the Kingman coalescent is a robust approximation to many underlying mating structures⁸, there are several mating schemes that may model the data inappropriately^{9,10}.

The key feature of the Kingman coalescent is that, at any given time, only a pair of individuals will share a common ancestor. This implies that parents have few offspring. However, in certain species, such as marine invertebrates, one individual may be responsible for a large fraction of the offspring in the next generation¹¹. Indeed, even some human individuals may have contributed highly disproportionately to offspring in future generations, such as Genghis Khan¹². In this case, more than two individuals may share a common ancestor at a time. Such models, in which more than two individuals can coalesce, are called lambda coalescents¹³. In other cases, a coalescent with multiple mergers may be used to model data in which the sample size is very close to or exceeds the effective population size, as is increasingly common in human genomics¹⁴.

While the theory of lambda coalescents is fairly young, there have been some attempts to apply them to inference in empirical data sets^{15–17}. Such inferences have suggested that it may be possible to learn from the data whether or not there are violations of a Kingman coalescent, and the site frequency spectrum may be a powerful tool to differentiate lambda coalescents from the standard coalescent in a complicated demographic scenario¹⁸.

Supplementary reference list

1. Ségurel, L., Wyman, M. J. & Przeworski, M. Determinants of Mutation Rate Variation in the Human Germline. <http://dx.doi.org/10.1146/annurev-genom-031714-125740> **15**, 47–70 (2014).
2. Kong, A. *et al.* Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**, 471–475 (2012).
3. O'Roak, B. J. *et al.* Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nature Genetics* **43**, 585–589 (2011).
4. Conrad, D. F., Keebler, J., DePristo, M. A. & Lindsay, S. J. Variation in genome-wide mutation rates within and between human families. *Nature* (2011).
5. Roach, J. C. *et al.* Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328**, 636–639 (2010).
6. Langergraber, K. E. *et al.* Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15716–15721 (2012).
7. Kingman, J. F. C. The coalescent. *Stochastic Processes and their Applications* **13**, 235–248 (1982).
8. Möhle, M. Robustness results for the coalescent. *J. Appl. Probab.* (1998). doi:10.2307/3215697
9. Sagitov, S. The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* **36**, 1116–1125 (1999).
10. Möhle, M. & Sagitov, S. A Classification of Coalescent Processes for Haploid Exchangeable Population Models. *The Annals of Probability* **29**, 1547–1562 (2001).
11. Eldon, B. & Wakeley, J. Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics* **172**, 2621–2633 (2006).
12. Zerjal, T. *et al.* The Genetic Legacy of the Mongols. *Am. J. Hum. Genet.* **72**, 717–721 (2003).
13. Pitman, J. Coalescents with multiple collisions. *Ann. Appl. Probab.* (1999). doi:10.2307/2652847
14. Wakeley, J. & Takahashi, T. Gene genealogies when the sample size exceeds the effective size of the population. *Mol Biol Evol* **20**, 208–213 (2003).
15. Birkner, M., Blath, J. & Steinrücken, M. Importance sampling for Lambda-coalescents in the infinitely many sites model. *Theor. Popul. Biol.* **79**, 155–173 (2011).
16. Eldon, B. Estimation of parameters in large offspring number models and ratios of coalescence times. *Theor. Popul. Biol.* **80**, 16–28 (2011).
17. Steinrücken, M., Birkner, M. & Blath, J. Analysis of DNA sequence variation within marine species using Beta-coalescents. *Theor. Popul. Biol.* **87**, 15–24 (2013).
18. Eldon, B., Birkner, M., Blath, J. & Freund, F. Can the site-frequency spectrum distinguish exponential population growth from multiple-merger coalescents? *Genetics* **199**, 841–856 (2015).